Aging and Engaging: A Social Conversational Skills Training Program for Older Adults

Mohammad Rafayet Ali[1], Kimberly Van Orden[2], Kimberly Parkhurst[3]
[1]Computer Science, University of Rochester
[2]University of Rochester School of Medicine
Rochester, United States.
[3]University of Utah.
Salt Lake City, United States.

Shuyang Liu[4], Viet-Duy Nguyen[5], Paul Duberstein[6], M. Ehsan Hoque[7]
[4, 5, 7]Computer Science, University of Rochester.
[6]University of Rochester School of Medicine.
Rochester, United States.

ABSTRACT

We developed "Aging and Engaging," a web-based intelligent interface, to improve communication skills among older adults. The interface allows users to practice conversations with a virtual assistant and receive feedback on eye contact, speaking volume, smiling, and valence of speech content. Feedback is generated automatically by analyzing the temporal properties of the conversation using the hidden Markov model. The interface was designed with the assistance of an expert advisory panel that works with geriatric patients, as well as a focus group of 12 older adults. To evaluate its effectiveness, we conducted a study with 25 older adults, each of whom participated in four conversations. Participants' response times to questions, as well as the amount of positive feedback, increased gradually through these interactions, as assessed by human judges. Participants found the feedback useful, easy to interpret, and fairly accurate, and expressed their interest in using the system at home. We plan to enroll subjects with difficulties in social communication; have them use the system over time at home in a randomized, controlled study; and measure any changes in their behavior.

Author Keywords
Aging, virtual agent, social skills.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION

Ed is 72 years old and lives alone. Recently, he began case management with his local aging services provider at the request of his primary care physician, given Ed's reports of social isolation and loneliness. The case manager visited

Figure 1. Aging and Engaging virtual conversation agent

Ed's home and immediately noticed that Ed rarely smiled and had a flat expression on his face. Ed reported that he had a few acquaintances, but no close friends. The case manager encouraged Ed to visit his local senior center. The manager was delighted to learn that Ed went to the center for lunch one day. However, Ed did not have a positive experience. He felt left out even among the group of older adults. He reported feeling uncomfortable about asking to join a card game. Instead, Ed sat by himself. One gentleman invited him to sit at his table at lunch, but Ed found he had little to say. After eating, the gentleman moved on to a different group. Ed didn't understand why he didn't seem to fit in, but the case manager noted to herself that Ed demonstrated difficulties with conversational skills. Specifically, Ed often failed to make eye contact with her, spoke in a monotone, and said little. When he did speak, he was dour and gloomy. If Ed interacted this way at the senior center, it would explain why he had trouble fitting in. The case manager correctly identified Ed's problem, but how could she help someone who was reluctant to talk to even a counselor?

People around the world are living longer—a phenomenon termed "population aging." According to U.S. Census Bureau projections, the worldwide population of older adults (those who are 65 years old or older) is projected to reach 1.5 billion by 2050. It is, therefore, a high priority to promote healthy (or "successful") aging. This is achievable. Later life is not typically a time of isolation and despair: despite losses

IUI'18, March 7–11, 2018, Tokyo, Japan © 2018 Association for Computing Machinery. ACM ISBN 978-1-4503-4945-1/18/03…$15.00
https://doi.org/10.1145/3172944.3172958

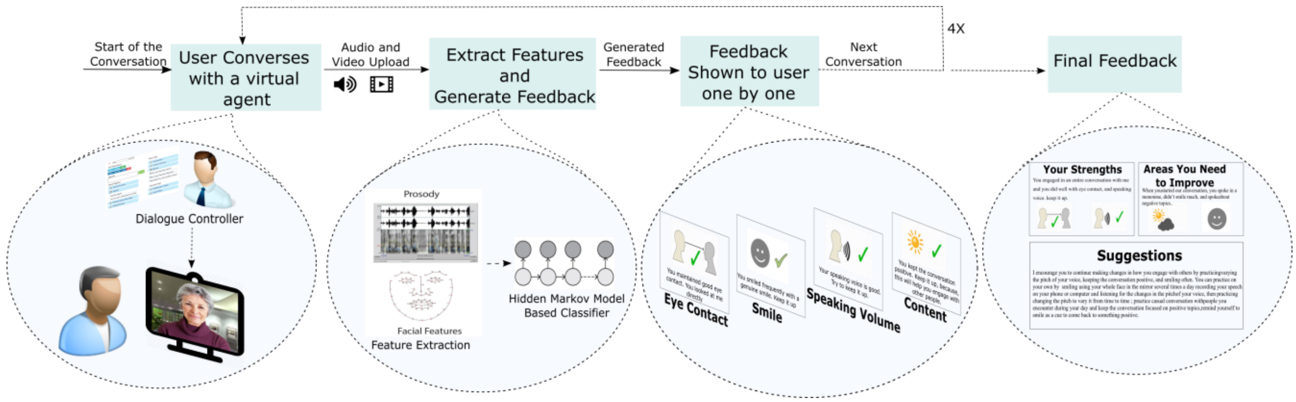


Figure 2. The overall system. Users initiate a conversation, which is driven by a dialogue controller. Conversation audio and video are processed in the server and feedback are generated. Users receive the feedback one by one and move to the next conversation phase. After four rounds of conversation, users receive final feedback summarizing the previous feedback.

in function, cognitive capacity, and social network size, rates of depression decline in later life [1]. Research indicates that some (but not all) negative emotions are experienced less often (e.g., anger) and that the ability to manage and change one's emotions (i.e., emotion regulation) often improves with age, in part due to increased motivation to experience positive emotions and accumulated life experiences that allow older adults to make effective choices regarding situations to approach or avoid [51]. Research from lifespan developmental theorists [2] has described the best approach to healthy aging as involving an increased focus on the most meaningful aspects of life—typically relationships—rather than a focus on losses or death. Some older adults, however, struggle with the transitions of later life and do become isolated and depressed [3]. A large body of literature has demonstrated that ineffective nonverbal communication impairs the ability to form and maintain positive, supportive relationships [6,7,8,9,10]. Up to 20 percent of community-dwelling older adults demonstrate difficulties with nonverbal communication [11], and yet effective intervention strategies are unavailable to them.

Some computer-based communication skills training programs exist [20,21,28,36,50]. These programs are difficult for older adults like Ed to use, for several reasons. They rely on a capacity for divided attention [36], which declines with normal cognitive aging. They require significant comfort and some experience with technology, which many older adults lack (though research has shown that they do well with technology when given appropriate instruction), and they do not accommodate for age-related sensory changes. In this paper, we propose a conversational skills training system geared toward older adults called "Aging and Engaging." In its initial application, it is designed for older adults who visit senior centers. Not all older adults who visit senior centers are able to have positive and meaningful social interactions, precisely because they lack the conversational skills to do so. Our interface allows them to converse with a virtual assistant and receive feedback on three nonverbal behaviors—eye contact, smiling, speaking volume—and one verbal: valence of speech content. We chose these four behaviors because they are empirically linked to poor social functioning [8,17]. Other communication behaviors, like gesturing, can be affected by physical impairments due to chronic disease. We divide each conversation into four phases, allowing users to receive and reflect on feedback often. In each phase, users interact with the assistant (see Figure 1) for about four minutes. During the conversation, our system captures audio and video and uploads both to our server in real time for immediate analysis. From the uploaded video and audio files, the facial and prosodic features, including smile intensity, volume, and eye gaze direction are extracted. A hidden Markov model–based classifier then classifies the patterns of nonverbal features into two categories: positive and negative. After each conversation phase, our system gives the feedback based on the classified temporal patterns.

Our system performs automated speech recognition, which we later use to perform a sentiment analysis (ratio of positive to negative words, for instance). After each phase, the transcript is uploaded to the server, where our system looks for negative and positive words in the transcript from a pre-populated list. The list of words was generated with direct input from two clinical psychologists who provide regular therapy to elderly patients. Our system gives overall positive or negative feedback depending on the prediction and on the three nonverbal and one verbal cue, after each phase. The four types of feedback come one after another with both text and voice options to improve accessibility. After the end of the conversation, our system generates a summary of all the feedback provided during each phase, identifying users' strong areas, as well as weaknesses. The feedback also includes suggestions for improvement. We employed a dialogue controller using a Wizard-of-Oz technique, which uses an online interface to select dialogues from predefined topics. Figure 2 shows the key components of our system. We aim to develop a system that engages users in conversation, puts at ease, and offers useful feedback that is interpretable.

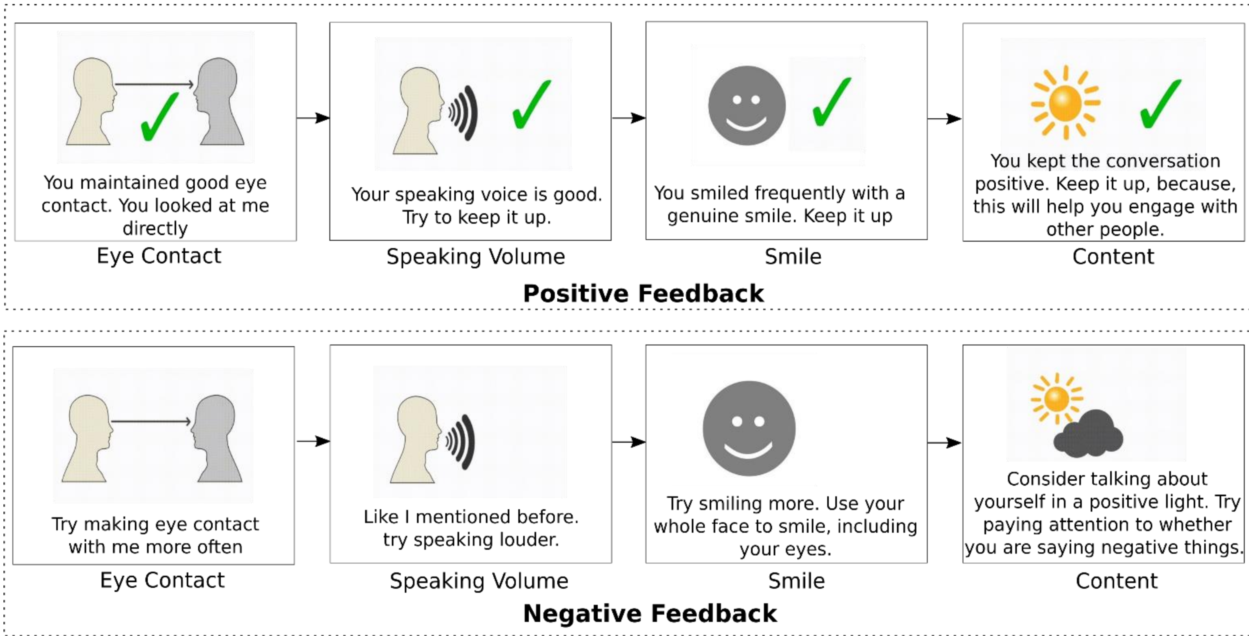


Figure 3. An example of feedback interface for each conversation phase. Users can receive either positive or negative feedback for each of the four conversational skills cues. For example, a user can receive positive feedback on eye contact and speaking volume, and negative feedback on smiling and content.

Our interdisciplinary team developed a semi-automated interface with the input of an expert advisory panel of professionals working with geriatric patients (geriatricians, neuropsychologists, gerontologists) and a focus group of 12 older adults. Our focus group helped in designing the interface and feedback, as well as selecting dialogue topics of the conversations. We conducted a study with 25 participants, each of whom is 60 years old or older. We collected the videos of their conversations with the virtual agent. Later, two raters watched the videos independently and determined whether the system should have provided positive or negative feedback. (Inter-reliability was 0.73.) Our system showed an accuracy of 72 percent, on average, considering the human-provided annotations as a ground truth. The post-study interviews show that our participants found this program useful, easy to interpret, and easy to use. (Further details are provided in the “Qualitative Analysis” section.) They also found, however, that the system’s feedback was sometimes inaccurate. Unsurprisingly, participants felt the conversation was not as natural as a face-to-face interaction with another human. According to human evaluations, our participants gradually received less negative feedback as they continued the session ($p < 0.05$) and increased their speaking time (Δmean = 15.79 sec, ΔSD = 4.43, $p < 0.05$). Using the video data and annotations collected, we intend to deploy a fully automated program for use at home in the near future.

RELATED WORK

Prior technologies focused on providing more opportunities for social interactions among older adults. Beacker et al. [18] designed a social media platform, InTouch, to help older adults share photos, text, audio, and video messages within their network. Garattini et al. [19] developed a system, the Building Bridges Device, to create social networks. This device contains a touchscreen computer and a phone cradle that allows users to chat, make phone calls, and broadcast or send asynchronous messages. Their study of 19 older adults over a 10-week span showed no significant decline in usage when paired with face-to-face meetings. That is, when the technology is paired with face-to-face interactions, the attrition rate does not change.

Numerous studies used virtual agents to provide companionship to older adults. The main goal of these projects was to explore different ways of reducing social isolation. This differs from our work, as we focus on improving social skills. Vardoulakis et al. [20], for example, designed a virtual agent companion. They used a Wizard-of-Oz technique to collect data from older adults who were socially isolated. Participants’ self-reported ratings showed that they were willing to use the device again. Ring et al. [21] developed a computer program with a virtual assistant that exhibits synchronous non-verbal behavior using the Behavior Expression Animation Toolkit (BEAT) [22]. Their program has two modes: passive and proactive. In proactive mode, the program can initiate a conversation after observing the user with motion sensors. In passive mode, users had to initiate the conversation by themselves. In a one-week study with 14 older adults, the proactive group reported a significantly greater decrease in loneliness than the passive group. Bickmore et al. [23] developed a virtual lab for conducting studies with human and virtual agents. They used it as an exercise coach for older adults. Participants who used

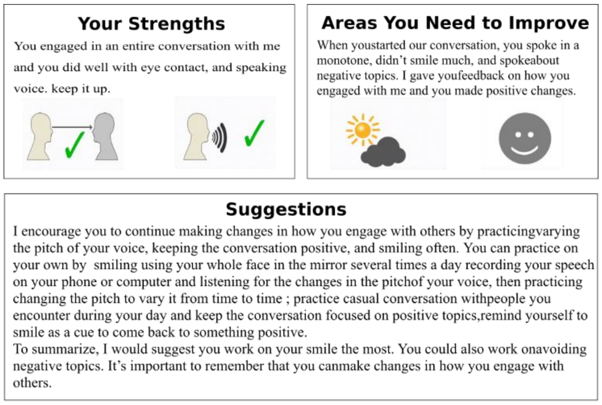


Figure 4. Final feedback interface

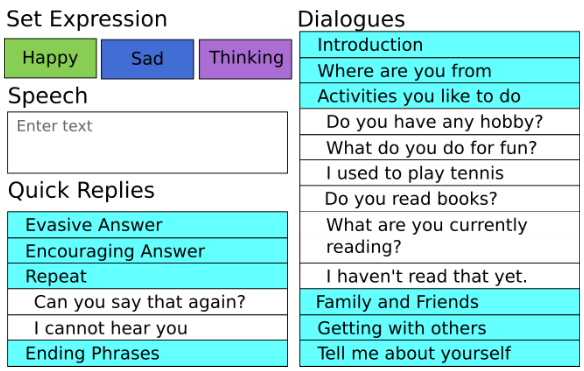


Figure 5: The dialogue controller's interface

the system exercised significantly more. Previous efforts with younger adults focused on giving real-time and *ex-post facto* feedback on nonverbal behavior. These efforts have not been replicated for older adults. Some of the previous studies focus on minimizing the distraction caused by real-time feedback [24,25,26]. (Older adults perform worse when their attention is divided [27].) In the past, *ex-post facto* summary feedback has been effective. In My Automated Conversation coacH (MACH) [28], a virtual agent asks a series of interview questions, providing neutral acknowledgments by mirroring smiles and head nods and then following up with detailed feedback on the user's performance. This project demonstrated the viability of using virtual agents and *ex-post facto* feedback in the context of a job interview. ROC Speak [29] is a semi-automated public speaking practice tool that uses crowdsourced and automated nonverbal behavior analysis. It, too, employed *ex-post facto* feedback, which enabled users to reflect on their mistakes and practice again.

We focus on conversational skills training to help older adults improve social communication. This is significantly different than previous studies [18,19], which focused on developing social networks and reducing social isolation by providing a virtual agent companion. Our system is a coach, not a companion. In the past, the social skills training tools have demonstrated the effectiveness of the automated computer-based social skills training. However, these tools do not focus on the older adults who have particular needs. For example, using MACH [28], the participants, who were college students showed significant improvement on job interviews. Aging and Engaging is inspired by the concept of using a virtual agent and adapted to help improve social skills among older adults. The interface does not feature any graph or numbers, instead, it uses pictures and text to speech technique to deliver the feedback. Unlike past interfaces [24,26,36,50] which aimed to change behaviors among young adults, our interface does not feature any real-time feedback. Our work draws on previous systems that aimed to change nonverbal behaviors [28] but is innovative in that it was designed to meet the unique needs of older adults. Older adults who differ from younger adults with respect to sensory and cognitive processing, including vision and hearing loss [30], as well as declines in working memory and difficulties with divided attention [27].

SYSTEM

In designing a system for older adults, we encountered several challenges. First, we needed to design an interface that is simple and requires minimal training to use. Second, we needed to generate feedback on the nonverbal behaviors of the users. Third, our system needed to give feedback that is generalizable and easy to understand. To understand these problems better, we consulted with a focus group of 12 older adults, each of whom volunteers as companions for isolated peers. They have valuable experience with older adults who are at an elevated risk of having nonverbal behavioral difficulties. We described our concept and showed them a low-fidelity prototype. The focus group offered thoughtful advice on the virtual assistant's appearance, feedback, and dialogue topics. In summary, the focus group made the following suggestions: the avatar should look like an older adult; the feedback needs to be less distracting (i.e., avoid real-time feedback and have feedback appear in sequence, not all at once); avoid too much negative feedback, as it might disengage users; and reinforce positives and demonstrate understanding. We then consulted with the expert advisory panel. This panel consists of a geriatrician, a geriatric neuropsychologist, and a gerontologist with expertise in interventions. The advisory panel made the following suggestions: include short assessment points throughout the intervention in which the virtual assistant can give participants feedback directly—as opposed to real-time feedback, delivered through text and images—and also foster transitions; and the avatar's face should be highly contrasted against the background so that even visually-impaired participants will have no difficulty seeing it. We tried to incorporate all suggestions. We divided the conversation into four phases to reduce cognitive load on users. Our system gives feedback after each phase. Given that increasing difficulty with divided attention is common with cognitive aging [27], many older adults may find it difficult to focus on a conversation and feedback simultaneously [24,33,26]. This may be especially true of older adults with social functioning impairments, like those for whom our program is designed [31]. Additionally, feedback across four stages allows users to adjust their behavior in successive conversations. Afterward, our system summarizes the feedback and provides recommendations for future practice as well as areas to emphasize later (e.g., "vary pitch while explaining something").

Interface Design

Our interface features an older woman from SitePal [32] (Figure 1). We choose this virtual assistant after consulting with our focus group. Also, using the virtual agents for social skills training has been shown to be successful in different contexts and user groups [28,36,50]. During the conversation with the virtual assistant, our system uploads the recorded audio and video to a server for processing. After the conversation, our system uploads the transcript of the conversation to the server to generate feedback. The system automatically takes users to the feedback page. We chose four different dimensions of conversational skills to give feedback: eye contact, volume, smile, and speech content. The attendant behaviors (smiling, volume modulation, making eye contact, etc.) have been shown to foster communication [16,34,35]. A future version of our system may include other verbal and nonverbal cues. The feedback was given using pictures, text, and voice. Figure 3 shows our feedback interface. The top row shows an example of positive feedback; the bottom row, negative feedback. The feedback interface contains a picture and some text. Some studies used pictures to give feedback effectively [36]. We added text feedback because feedback offered through multiple channels is more likely to be effective than that offered through only one channel. We selected this approach in response to the expert panel and the focus group. For each skill dimension, our system offers positive and negative feedback. If the user makes eye contact with the avatar, for example, they receive positive feedback on eye contact. Each instance of positive feedback is accompanied by an animated green checkmark and a "ding" sound. The feedback text is read aloud to the user using a text-to-speech engine. To avoid repetition, the negative feedback text changes across each conversation. For example, if a user does not smile at all during the first conversation, the system will say, "You didn't smile at all. Try smiling more, including with your eyes." If the user does not improve in the second conversation, the system will say, "Like I mentioned before, try smiling more." To show engagement, we programmed our virtual assistant to nod its head periodically (every five seconds).

After every phase of the conversation, our system offers final, integrated feedback. This final feedback summarizes the feedback the user received after each phase. Figure 4 shows the final feedback interface. The final feedback shows, at most, two dimensions in which the user received negative feedback. We decided to show only two cases, as too much negative feedback can lead users to disengage. The system suggests ways for users to improve. For example, if the user receives negative feedback on speech content, our system will say: "Practice casual conversation with people you encounter during your day, and keep the conversation focused on positive topics."

Dialogue and Feedback Generation

Automated understanding of language and responding remains an active area of exploration. In this exploratory study, we decided to go with a Wizard-of-Oz technique to manage the dialogues. A research assistant worked as a dialogue controller. The dialogue controller uses an online interface shown in Figure 5. To avoid delay due to typing, we decided to generate dialogues on predefined topics and list the sentences in the dialogue controller's interface. On the right side of the interface (see Figure 5), we listed all dialogue options and grouped them by topic. After discussing with the focus group, we selected five topics: where the user is from, the activities the user likes to do, family and friends, getting together with others, and the user him or herself. In our study, the first topic (an introduction) was mandatory for all participants, after which we allowed participants to choose any three among the five topics for the remaining three phases. This is how we ensured that each of the participants completed four phases of conversation. The focus group helped us generate the prompts and replies for each topic. On the left side of the controller's interface (see Figure 5), there are buttons for controlling the assistant's nonverbal behavior, including selected dialogue topics, a few quick replies, and freeform text input.

The system records both audio and video of the conversation using the computer's microphone and webcam then sends the stream to our server in real time. On the server, the facial and prosodic features are extracted from the audio and video files using Praat [37] and SHORE [38]. Our features include smile intensity, pitch, volume, frequencies of the first three formants (F1, F2, F3), and pixel differences between consecutive frames. We used a face tracker from Visage technologies [39] to extract gaze direction in real time. To generate feedback, we used a hidden Markov model–based technique. In the past, hidden Markov model was used to model human behaviors and actions [55,56]. The training data was collected from a speed-dating study [36]. The data includes 506 minutes of video from 47 male college students having a conversation with a virtual assistant. The videos contain only the male students' faces. A trained set of research assistants from psychology with experience in social skills training watched the videos to mark those moments where they felt a subject needed to adjust their nonverbal behavior. Specifically, the research assistants looked for appropriate eye contact, volume, and smile. We extracted the features from the training videos and trained three hidden Markov models using the Baum-Welch algorithm [54] for the three nonverbal cues. To predict the feedback (positive or negative), we used the forward-backward algorithm and the features we extracted from participants' videos. Given a sequence of observation, the forward-backward algorithm calculates the posterior probability of a hidden Markov model [58]. To give feedback on content, our clinical psychologists collected some keywords, which we considered negative words in the context of our study (e.g. "died," "lonely"). We consider it as a negative conversation if the ratio of negative words and total words is greater than a predefined threshold empirically chosen after testing it on our research assistants. An example of false positive feedback would be, participant kept the conversation positive but the system gave a negative feedback on the content. A limitation of this approach is that it is unable to detect when a negative word is used in a positive way, and vice-versa. Adding this feature remains part of our future exploration.

STUDY

To assess the acceptability of the interface to its intended audience, we conducted a study with 25 participants who are each 65 years old or older. The average age of the participants was 68.5. Of those 25, six were single, eight were divorced, one was widowed, and ten lived with their spouses. Participants were recruited from a hospital-based geriatric mental health clinic (n=5) (where difficulties with nonverbal behaviors are common) as well as through print advertisements (n=20). All participants were native English speakers. Our participants varied on self-perceived social skills, as we believed more diverse feedback would be most useful at this stage. However, the majority of our participants (67%) were below the 50^{th} percentile on a population-normed measure of social skills (PROMIS Self-Efficacy to Manage Social Interactions). The study took place in a private room in the medical center. Participants first consented to the study with a trained assessor. We emphasized that participation would have no impact on services received from the Medical Center if any, and our research staff had no connection with the clinic. After the consent process, we explained how the interface works. Participants then interacted with our interface by pressing the start button in our interface. During the interaction, participants went through four conversation phases. After each phase, the system redirected participants to a feedback page. After four rounds of conversation and feedback, the participant received the final feedback (Figure 4), which summarized the feedback previously given and suggested areas for improvement. Our research staff did not intervene beyond helping participants open the interface in the web browser. After the interaction, we asked participants a series of questions about the interface, their experience, their demographics, self-perceived social skills (PROMIS) [52,53] any depression or anxiety [44], their social connectedness [45], and their use of computers or other electronics. The study served to collect data since there is no dataset available from any previous studies that target the desired behaviors in older adults. We did not conduct this study as a system intervention. We plan to conduct such an intervention study in the future to evaluate the effectiveness of the system.

Survey Results

We showed the participants four statements and asked them to rate each between one and five, where one indicates strong agreement and five indicates strong disagreement. The statements and their average ratings are shown in Table 1. On average, participants self-reported the interface as easy to use, with a value of 1.96 (SD=1.02). Our intuitive design choices, with minimal button-pressing and voice-assisted feedback, might have made our interface easy to use. Participants also disagreed with the statement about the system taking too long to use, rating it a 4.04 (SD=0.97). We wanted to know more about the opinions of the participants who agreed with the statement, "I feel disconnected from other people" (n=12). Table 1 shows no significant difference in ratings of the interface between participants who felt disconnected and those who did not. There was also no difference in the ratings provided by those who felt like outsiders at social events (n=12) and those who felt lonely (n=20). The participants were further divided based on a self-reported measure of perceived social skills (in Table 2). We did not find any significant difference in the ratings between the two groups. These results are important because they show the interface is acceptable to people who have (or are at risk for) difficulties with social communication [46].

In our study, 19 participants possessed a home computer or a laptop—we called them the "computer user group"—and six participants had no access to computers—the "computer non-user group." These two groups' ratings are also shown in Table 1. The computer non-user group thought the program was easier to use than the computer user group. This may be because the computer user group had higher expectations in terms of feedback speed and system responsiveness. The computer non-user group tended toward not wanting to use the program in their home. But these differences are not statistically significant. The lack of difference between computer users and computer non-users is important because it shows that our system is acceptable to even those who have limited experience with technology. We also looked at the average ratings of the statements after grouping participants based on the presence (or absence) of clinically-significant depression, anxiety, and social isolation, as these are characteristics that often coexist with communication difficulties in later life. Table 2 shows the average ratings. In it, the "yes" column contains the ratings of those who reported elevated levels of depression, anxiety, or social isolation using population-based norms. In these three categories, we did not find any significant differences in ratings between the groups, which, while inconclusive given our sample size, nonetheless suggests that those who feel isolated, depressed, or anxious are not less likely to find the program acceptable. This is important, as those who report these characteristics are most likely to have deficits that could improve the program.

To see how the system performed in terms of feedback accuracy, we compared the system-generated feedback against that of human judges. Two trained research assistants watched each of the interactions individually and decided whether to offer positive or negative feedback. We calculated the accuracy of our system by comparing against these determinations as ground truths. Table 3 shows the average accuracy of the system when compared against the decisions of the two human judges. The inter-rater reliability (Cohen's Kappa) of the two human judges for each feedback dimension is also shown in Table 3. We found that our system-generated feedback accorded most with human judges on eye contact. Some of our participants were not sure where to look to receive positive feedback. Others sat facing the light, which reflected on their eyeglasses, making it difficult to accurately determine eye gaze direction. Some dialogue topics became negative even when participants tried to keep them positive. Our system only looked at the ratio of positive words to negative words, not more complex syntactic structures. There were instances in which participants described a positive phenomenon with negative words and thus received negative feedback on their speech content.

Feature Analysis

We looked at how participants were performing in each phase of their conversation. We took the percentage of participants who received positive feedback according to our human annotators (see Figure. 6a). We saw a statistically significant difference between the first phase of conversation and the last (fourth) phase of conversation ($p<0.05$) on a smile, volume, and content. This suggests that participants were able to reflect on the feedback and change those behaviors more likely to elicit positive feedback. Among the four dimensions, the percentage of participants receiving positive feedback on eye contact changed the least (Δ = 0.31%), whereas smile changed the most (Δ = 83.15%).

We also looked at the average speaking time of the participants during each phase of the conversation (Figure. 6b). Participants' response times increased in subsequent conversations. Average response time for each question during the first conversation was 24 seconds (SD=4.5), whereas the average response time for the final conversation was 40 seconds (SD=4.9) ($p<0.05$). This suggests that, as the conversation progressed, the participants become more comfortable with the interface and revealed more information about themselves.

We further analyzed the affective features of the participants. We extracted the facial action units (AU) [47] features from the participants' videos using OpenFace [48] software. Each of the AU is expressed in two forms, a strength value ("_r") and binary presence ("_c"). Figure 6c shows the difference between the features of the first and last conversation phase where the difference was significant after adjusting for the number of features ($p<0.05$). We observed that the AU01 (inner brow raise) which is associated with a surprised facial expression [49], had an increase in both average and variance in the last conversation phase. In our context, this change may indicate that the participants were becoming more animated as the interaction progressed. AU04 (brow lower), the variance of AU09 (nose wrinkler), and variance of AU15 (lip corner depressor) also decreased in the last interaction. AU04, AU09, and AU15 are associated with the expressions of anger, disgust, and sadness respectively [49]. This might indicate that as the conversation went the participants became more comfortable.

Post Study Interviews

In order get additional feedback to help us improve our system, we included a qualitative interview with our participants (this interview was added midway into data collection, thus we only have feedback from n=15 subjects). We took a phenomenology approach to designing the interview guide, which was developed by one of our study investigators. Feedback was obtained from an additional researcher and the interview guide was edited accordingly. The interview guide included questions focused on participants' experience during the time they engaged with the program and suggestions for improving the program.

	Avg. Rating					
Statements	**AVG (n=25)**	**Disconnected (n=12)**	**Outsider (n=12)**	**Lonely (n=20)**	**Computer User(n=19)**	**Computer Non-User(n=6)**
1. The program was easy to use.	1.96	1.83	2.08	1.78	2.08	1.69
2. Using the program was frustrating.	3.68	3.67	3.83	3.94	3.69	3.67
3. The program took too long to use.	4.04	4.00	4.25	3.89	4.14	3.91
4. I would likely use the program if I could access it from home on a computer.	2.68	2.75	2.58	2.72	2.57	3.33

Table 1: Average ratings from participants in different groups based on interpersonal needs. (1 = strongly agree, 5=strongly disagree)

	Avg. Rating							
Statements	**Depression**		**Anxiety**		**Social isolation**		**Social Skills**	
	Y(n=1)	No	Y(n=3)	No	Y(n=2)	No	Good(n=6)	Poor
1. The program was easy to use.	1.00	2.00	1.67	2.00	2.00	1.95	1.78	2.11
2. Using the program was frustrating.	4.00	4.00	4.33	3.76	3.50	3.77	3.89	3.44
3. The program took too long to use.	4.00	4.04	3.67	4.09	4.50	4.00	4.22	3.68
4. I would likely use the program if I could access it from home on a computer.	3.00	2.70	3.60	2.81	3.00	2.68	2.67	2.44

Table 2: Average ratings from participants in different groups based on their diagnosis (1=strongly agree, 5=strongly disagree

	Accuracy (%)	True Positive Rate (%)	False Positive Rate (%)	Cohen's *kappa*
Eye Contact	79.60	76.06	3.50	0.84
Volume	72.00	58.50	14.25	0.79
Smile	66.00	60.25	15.35	0.62
Content	68.80	67.91	17.36	0.67

Table 3: Accuracy of feedback compared to human judges

Interviews were audio-recorded with participants' permission and transcribed. We then conducted thematic analysis using principles of grounded theory (including the constant comparative method) to analyze the data [57]. Coding was done by one investigator and then reviewed by another investigator (we used open coding) [57]. Discrepancies between the two raters did not arise as the themes were relatively straightforward and unambiguous. We added user codes with the quotations.

Accuracy of the feedback

All participants commented on their perceptions of the accuracy of the feedback, as they reflected on their performance and suggested areas for improvement. Many participants perceived that the feedback was accurate (n=5).

"Every single thing she said I should try to be better at was absolutely correct." (Subject 2, 69 y/o female, 45th percentile on social skills).

Insight into their behaviors and its relation to feedback accuracy was also raised by several participants (n=3).

"I'm pretty much aware of all those things I don't do well. I don't smile a lot, I don't vary the pitch of my voice. I have a very soft voice." (Subject 14, 74 y/o female, 39th percentile for social skills).

However, an equal number of participants thought the feedback was inaccurate (n=5). This may be the case because the training data was collected from college students, rather than older adults. A key objective of the current study was to obtain data from older participants to better train the system for our subsequent efficacy study. One way in which our older participants differed from many of the younger participants was that most wore eyeglasses, which made it difficult for the program to recognize eye gaze direction. Further, some of the participants were not sure where to look (at the webcam or into the virtual assistant's eyes) in order to get positive eye contact feedback. At first, this frustrated them, but eventually, they understood and looked at the assistant's eye to get a positive feedback. As another said,

"I was frustrated by inaccurate feedback, but once it seemed to be in sync, it was fine."(Subject 11, 70 y/o female, 18th percentile on social skills).

Comfort with the interface

Most participants (n=10) spoke about their level of comfort with the interface given that it was a novel experience for all of them. Some spoke about how they initially felt uncomfortable engaging with the program, but that this improved over the course of the session.

"I got more comfortable as I went on because I wasn't sure how it was going to work at first." (Subject 2, 69 y/o female, 45th percentile on social skills). Another said, "'I think the more it went on, the more comfortable you get with it." (Subject 7, 69 y/o male, 55th percentile on social skills) Many participants also spoke about ways in which the program was easy or difficult to use (n=8). We carefully designed our system using the feedback from our focus group. We had several key design concerns, which included participants' ability to read the feedback, navigate through the interface, and hear properly. We minimized button pressing to make the program more intuitive. Said another:

"The program is easy to use; I didn't need training like I do with other new programs." (Subject 11, 70 y/o female, 18th percentile on social skills).

Useful program for engaging

We asked our participants about how useful they thought the system was. Most participants thought the program could help improve conversational skills (n=12). Our interface allows these people to practice and possibly improve conversational skills without feeling judged by a human. Said one of our participants:

"I know that I could probably improve by practicing. I've never had the opportunity to practice with something like this, which would be an impersonal coach. I started to think of her as a coach who wouldn't react—like a psychiatrist, just listening and letting the person do all the talking." (Subject 14, 74 y/o female, 39th percentile for social skills).

Our participants also said the program could be a great teaching tool. Said another,

"I think it's a great teaching tool. The program is already helpful; it just needs tweaking." (Subject 3, 67 y/o female, 32nd percentile on social skills).

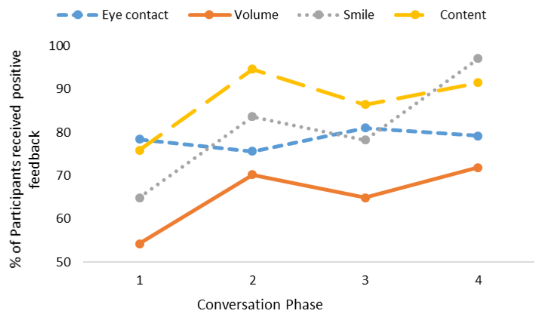
Our participants also viewed this program as a tool that can be helpful to those who are shy and withdrawn, learning the skills needed to connect and communicate with others. This program might also be helpful for those who find it difficult to start a conversation.

"This would be good for people who have issues—who are so withdrawn that they can't communicate. This would help them break out of their shell, and give them instructions on how to move forward." (Subject 8, 65 y/o female, 70th percentile on social skills).

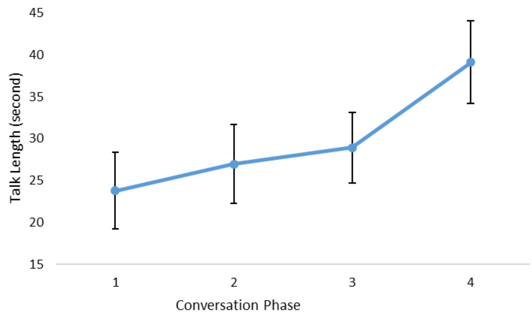
Content of the discussion

We carefully chose dialogue topics after discussing with our focus group. Several participants (n=5) commented on the content of the dialogue.

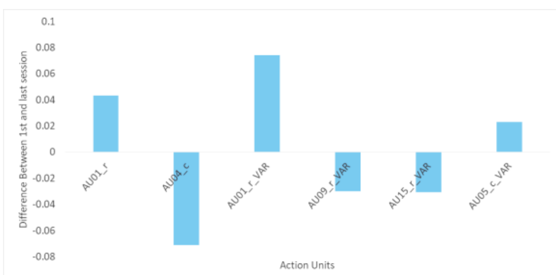
"I would've done all five conversation topics- they were great. There was nothing intrusive, it was very basic."



(a)



(b)



(c)

Figure 6. (a) The trend line of the percentage of participants annotated as positive feedback recipients after each conversation. (b) The average length of the participants' speeches for each of the small conversation. (c) Changes in the average action unit values.

(Subject 18, 62 y/o female, 25th percentile on social skills).

During the conversations, the virtual assistant asked some questions relevant to their selected dialogue topics. Our participants found that the questions the virtual assistant asked were common, natural, and comprehensive.

"I thought the questions were so realistic. Those are the questions I grill people on all the time when I meet them. It was very natural." (Subject 8, 65 y/o female, 70th percentile on social skills)

Some of the participants felt that the conversation was one sided, as the virtual assistant was not able to respond to all the questions the participant asked.

"I gave answers but I didn't feel like I had the chance to ask her questions. It felt strange... The conversation is too one-sided." (Subject 2, 69 y/o female, 45th percentile on social skills)

Suggestions to improve the interface

Two participants addressed the fact that the conversations had no time limit, allowing participants to speak as long as they wanted. It was not clear to the participants how long the conversations would continue. One participant suggested using a timeline.

"A timeline would be helpful—knowing how long each section is going to be, and how far along you are in the program." (Subject 5, 83 y/o male, 61st percentile on social skills).

Another participant brought up the issue of sensory impairment in later life.

"At first I had a hard time hearing, but after she started talking, you kind of get what was going on. It could be part of the hearing problems I have, but once I got into it, I could understand her voice. As it went on it got better." (Subject 7, 69 y/o male, 55th percentile on social skills)

To further improve our program, we plan to consult with the geriatrician on our advisory board for features we could add to better accommodate age-related hearing and vision loss.

DISCUSSION AND FUTURE WORK

Even though the agreement between the feedback generated by our algorithm and two labelers was fairly high, there is room for improvement. The feedback module was trained with videos of college students who are younger than our target population and have different facial, prosodic and behavioral characteristics. We used a hidden Markov model to better understand temporal patterns to generate feedback. In future, we will continue to improve our proposed machine learning model to generate more accurate feedback. Our exploration will include other non-temporal models (i.e., support vector machine) and neural network based classifiers (i.e., recurrent neural network). In this exploratory study, we collected 25 videos of interaction with older adults. We will improve our feedback module by using this new set of videos. Formulating feedback by counting positive and negative words has many limitations. Future work will involve more semantic analysis, using techniques from natural language processing.

The effectiveness of the system (i.e., improving social skills) was not assessed through the preliminary study. The objective of the study was to assess the acceptability of the system in older adults. This type of intervention is acceptable to younger adults. There are no evidence-based reasons to predict that older adults would react differently than younger adults to the prospect of using the program. Some of the end users take part in the design of the system. Some of the older adults we included did have difficulties with social skills. We also included those without these difficulties because we believe a wider range of feedback is helpful. In future, we will recruit participants with social skills deficits.

The study was a one-time visit by the participants. Thus, it was not necessary to remember the information from the previous sessions, and there was none. However, the system remembers the feedback it gave to the participant in subsequent conversations (in the same day) and synthesizes them to produce a final feedback. In future, we will enable the system to remember the past conversation by augmenting the login capability.

Participants appreciated the dialogue we chose. The interactions, however, were one-sided, as the virtual assistant was not able to answer all the questions posed by our participants because we generated a limited number of responses. The transcripts collected from this study will allow us to generate more responses for the assistant. To automate the dialogue controller, we will build a topic-based dialogue module. The automated dialogue module can have conversations on particular topics and the responses will be chosen by pattern-matching. Spontaneous back-channeling was not implemented in our current prototype. In the future, we will incorporate features, such as sharing a smile or nodding appropriately, which can lead to a more natural interaction. This remains an open research problem for us and the virtual assistant community.

Our aim is to build an interface that will become a part of the users' lives. With the initial exploratory study and findings, our next step is to deploy a fully automated system online. We used the Wizard-of-Oz technique as our first step to build the automated dialogue system. This study gave us the valuable data necessary for an automated dialogue manager. It will allow us to run a longitudinal study allowing participants to use our system from home. It is important to investigate how the participants are able to utilize their practice with our interface to improve or maintain their relationships and maintain better health and quality of life. We will have older adults who demonstrate difficulties with social communication (as measured by low levels of social functioning and impairments on standardized ratings of communication behaviors) engage with the program several times and assess their communication behaviors at baseline, at the end of the two weeks of practice, and three months later. Doing so will allow us to measure learning and change over time, as well as maintenance of gains after finishing their sessions. Our study presented here is a necessary first step to this future study.

Further, later life brings with it numerous potentially stressful social encounters, including discussions about end-of-life care, for example. Even older adults with strong conversation skills will likely to have some difficulties with these conversations. Thus, a future direction with our system is to adapt it for assisting all older adults with challenging social situations common in later life, including end-of-life planning, discussions with physicians, discussions with adult children about giving up driving and moving to retirement communities.

CONCLUSION

Our semi-automated interface was designed to benefit older adults who have difficulties with communication skills, including those who are homebound, while avoiding social stigma and minimizing burdens associated with in-person interventions. Since our target users are different than those studied using previous social skills training tools, we spoke with the advisory panel and the focus group in order to make the design choices. Our exploratory study, with 25 participants, revealed key challenges while also underscoring the acceptability and feasibility of the system. Our participants shared their experiences using the interface and, overall, found it useful, intuitive, and fairly accurate. This indicates that such technology might be accepted by many older adults, including those who could stand to benefit the most. Our analysis of the response time and the human annotations revealed that participants gradually spoke with the assistant longer and received less negative feedback as the interaction progressed. The exploratory study provided important insights that we will use in building the next version of this program.

Improving interpersonal skills is a personal experience and should happen with the presence of other humans. Thus, our program is not designed to replace human interaction and companionship. Rather, the program is a tool to be used to improve communication skills in a safe, non-judgmental environment; these skills can then be practiced with other people. Having difficulties with conversation skills can make social interactions stressful, which exacerbates cognitive changes associated with aging that could impede learning new behaviors; using our program allows older adults to practice new skills in a less stressful environment before attempting to use new behaviors with people in their lives. Further, later life imposes a unique challenge for individuals who are homebound and unable to take advantage of help by commuting to a clinic, thus indicating an important role for a web-based automated program. Our program may help these individuals improve their skills, thereby maximizing the time they have with people in their lives and promoting useful and less stressful conversations with spouses and family members. We believe computation has an important role to play by providing fully automated, repeatable, standardized, and non-judgmental "conversational coaching" to individuals who may need it.

ACKNOWLEDGEMENT

The work was partially supported by NSF EAGER grant IIS-1543758 and by a University Research Award from the University of Rochester (Van Orden, PI).

REFERENCES

1. S. T. Charles, "Strength and vulnerability integration: A model of emotional well-being across adulthood." Psychol. Bull., vol. 136, no. 6, pp. 1068–1091, 2010.
2. S. T. Charles and L. L. Carstensen, "Social and emotional aging." Annu. Rev. Psychol., vol. 61, pp. 383–409, 2010.

3. K. A. Van Orden, A. A. O'Riley, A. Simning, C. Podgorski, T. M. Richardson, and Y. Conwell, "Passive suicide ideation: An indicator of risk among older adults seeking aging services?," Gerontologist, vol. 55, no. 6, pp. 972–980, 2015.
4. J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, "Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review," Perspect. Psychol. Sci., vol. 10, no. 2, pp. 227–237, 2015.
5. A. E. Schneider, N. Ralph, C. Olson, A.-M. Flatley, and L. Thorpe, "Predictors of Senior Center Use among Older Adults in New York City Public Housing," J. URBAN Heal. NEW YORK Acad. Med., vol. 91, no. 6, pp. 1033–1047, 2014.
6. R. Schwartz and M. D. Pell, "When emotion and expression diverge: The social costs of Parkinson's disease," J. Clin. Exp. Neuropsychol., vol. 3395, no. August, pp. 1–20, 2016.
7. C. F. Keating, "The developmental arc of nonverbal communication: Capacity and consequence for human social bonds.," in APA handbook of nonverbal communication., 2016, pp. 103–138.
8. M. A. Struchen, M. R. Pappadis, A. M. Sander, C. S. Burrows, and K. A. Myszka, "Examining the contribution of social communication abilities and affective/behavioral functioning to social integration outcomes for adults with traumatic brain injury," J. Head Trauma Rehabil., vol. 26, no. 1, pp. 30–42, 2011.
9. J. K. Burgoon, laura K. Guerrero, and V. Manusov, "Nonverbal signals," in Handbook of Interpersonal Communication (4th ed.), 2012, pp. 239–280.
10. J. K. Burgoon and A. E. Bacue, "Nonverbal communication skills.," in Handbook of Communication and Social Interaction Skills, 2003, pp. 179–219.
11. H. M. Geurts, M. Stek, and H. Comijs, "Autism characteristics in older adults with depressive disorders," Am. J. Geriatr. Psychiatry, vol. 24, no. 2, pp. 161–169, 2016.
12. A. S. Young, R. Klap, C. D. Sherbourne, and K. B. Wells, "The quality of care for depressive and anxiety disorders in the United States," Arch Gen Psychiatry, vol. 58, no. 1, pp. 55–61, 2001.
13. M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing acceptance of assistive social agent technology by older adults: The almere model," Int. J. Soc. Robot., vol. 2, no. 4, pp. 361–375, 2010.
14. K. Laguna and R. L. Babcock, "Computer anxiety in young and older adults: Implications for human-computer interactions in older populations," Comput. Human Behav., vol. 13, no. 3, pp. 317–326, 1997.
15. T. L. Mitzner, J. B. Boron, C. B. Fausset, A. E. Adams, N. Charness, S. J. Czaja, K. Dijkstra, A. D. Fisk, W. A. Rogers, and J. Sharit, "Older adults talk technology: Technology usage and attitudes," Comput. Human Behav., vol. 26, no. 6, pp. 1710–1721, 2010.
16. C. R. Bowie, W. W. Leung, A. Reichenberg, M. M. McClure, T. L. Patterson, R. K. Heaton, and P. D. Harvey, "Predicting Schizophrenia Patients' Real-World Behavior with Specific Neuropsychological and Functional Capacity Measures," Biol. Psychiatry, vol. 63, no. 5, pp. 505–511, 2008.
17. M. Youngren and P. M. Lewinsohn, "The functional relation between depression and problematic interpersonal behavior.," J. Abnorm. Psychol., vol. 89, no. 3, pp. 333–341, 1980.
18. R. Beacker, K. Sellen, S. Crosskey, V. Boscart, and B. Barbosa Neves, "Technology to reduce social isolation and loneliness," Proc. 16th Int. ACM SIGACCESS Conf. Comput. Access. - ASSETS '14, no. November 2015, pp. 27–34, 2014.
19. C. Garattini, J. Wherton, and D. Prendergast, "Linking the lonely: An exploration of a communication technology designed to support social interaction among older adults," Univers. Access Inf. Soc., vol. 11, no. 2, pp. 211–222, 2012.
20. L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, and T. Bickmore, "Designing relational agents as long term social companions for older adults," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7502 LNAI, pp. 289–302, 2012.
21. L. Ring, B. Barry, K. Totzke, and T. Bickmore, "Addressing loneliness and isolation in older adults: Proactive affective agents provide better support," Proc. - 2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact. ACII 2013, pp. 61–66, 2013.
22. J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, "BEAT: the Behavior Expression Animation Toolkit," Life-Like Characters, vol. 137, no. August, pp. 163–185, 2004.
23. T. Bickmore and D. Schulman, "A virtual laboratory for studying long-term relationships between humans and virtual agents," … Conf. Auton. Agents …, pp. 297–304, 2009.
24. O. Kulyk, J. Wang, and J. Terken, "Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings," in Lecture Notes in Computer Science, 2006, vol. 3869 LNCS, pp. 150–161.
25. T. Kim, A. Chang, L. Holland, and A. S. Pentland, "Meeting mediator: enhancing group collaborationusing sociometric feedback," in Proceedings of the 2008 ACM conference on Computer supported cooperative work, 2008, pp. 457–466.
26. M. I. Tanveer, E. Lin, and M. E. Hoque, "Rhema : A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking," in IUI 2015, pp. 286–295.
27. S. Getzmann, E. J. Golob, and E. Wascher, "Focused and divided attention in a simulated cocktail-party situation: ERP evidence from younger and older adults," Neurobiol. Aging, vol. 41, pp. 138–149, 2016.
28. M. (Ehsan) Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated

conversation coach," Proc. 2013 ACM Int. Jt. Conf. Pervasive ubiquitous Comput. - UbiComp '13, p. 697, 2013.
29. M. Fung, Y. Jin, R. Zhao, and M. E. Hoque, "ROC Speak: Semi-Automated Personalized Feedback on Nonverbal Behavior from Recorded Videos," Proc. 17th Int. Conf. Ubiquitous Comput., 2015.
30. C. Correia, K. J. Lopez, K. E. Wroblewski, M. H. Scheetz, D. W. Kern, R. C. Chen, L. P. Schumm, W. Dale, M. K. McClintock, and J. M. Pinto, "Global sensory impairment in older adults in the United States," J. Am. Geriatr. Soc., vol. 64, no. 2, pp. 306–313, 2016.
31. J. T. Cacioppo and L. C. Hawkley, "Perceived social isolation and cognition," Trends Cogn. Sci., vol. 13, no. 10, pp. 447–454, 2009.
32. "Site Pal." [Online]. Available: www.sitepal.com.
33. G. Leshed, D. Perez, J. T. Hancock, D. Cosley, J. Birnholtz, S. Lee, P. L. McLeod, and G. Gay, "Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups," Proc. 27th Int. Conf. Hum. factors Comput. Syst. CHI 09, p. 537, 2009.
34. J. L. Hames, C. R. Hagan, and T. E. Joiner, "Interpersonal processes in depression.," Annu. Rev. Clin. Psychol., vol. 9, pp. 355–77, 2013.
35. C. Segrin, "Social skills deficits associated with depression," Clin. Psychol. Rev., vol. 20, no. 3, pp. 379–403, 2000.
36. M. Ali, D. Crasta, L. Jin, A. Baretto, J. Pachter, R. Rogge, and M. (Ehsan) Hoque, "LISSA- Live Interactive Social Skill Assistance," in Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, 2015, pp. 173–179.
37. P. Boersma and D. Weenink, "Praat: doing phonetics by computer." .
38. "SHORETM - Object and Face Recognition." [Online]. Available: http://www.iis.fraunhofer.de/en/ff/bsy/tech/bildanalyse/shore-gesichtsdetektion.html.
39. "Visage Technologies face tracking and analysis." [Online]. Available: http://visagetechnologies.com/.
40. An assessment of Interpersonal Competence, "Conversational Skills Rating Scale," 2007.
41. E. J. Finkel, P. W. Eastwick, and J. Matthews, "Speed-dating as an invaluable tool for studying romantic attraction: A methodological primer," Pers. Relatsh., vol. 14, no. 1, pp. 149–166, 2007.
42. D. Wendler, Improve your social skills. 2014.
43. "5 Body Language Tricks To Make Anyone Instantly Like You - Personality Development & English Lessons." .
44. W. T. Riley, P. Pilkonis, and D. Cella, "Application of the National Institutes of Health Patient-reported Outcome Measurement Information System (PROMIS) to mental health research," J. Ment. Health Policy Econ., vol. 14, no. 4, pp. 201–208, 2011.
45. K. A. Van Orden, K. C. Cukrowicz, T. K. Witte, and T. E. Joiner, "Thwarted belongingness and perceived burdensomeness: construct validity and psychometric properties of the Interpersonal Needs Questionnaire.," Psychol. Assess., vol. 24, no. 1, pp. 197–215, 2012.
46. A. Struchen, N. Clark, M. Sander, R. Mills, G. Evans, and D. Kurtz, "Relation of executive functioning and social communication measures to functional outcomes following traumatic brain injury.," NeuroRehabilitation, vol. 23, no. 2, p. 185, 2008.
47. P. Ekman and W. V Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. 1978.
48. T. Baltruˇ and P. Robinson, "OpenFace:an open source facial behavior analysis toolkit."
49. S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," Proc. Natl. Acad. Sci., 2014.
50. Anderson, Keith, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian et al. "The TARDIS framework: intelligent virtual agents for social coaching in job interviews." In Advances in computer entertainment, pp. 476-491. Springer, 2013.
51. Isaacowitz, D. M., Livingstone, K. M., & Castro, V. L. (2017). Aging and emotions: experience, regulation, and perception. Current opinion in psychology, 17, 79-83.
52. Hahn, E. A., et al. (2010). "Measuring social health in the patient-reported outcomes measurement information system (PROMIS): item bank development and testing." Qual Life Res 19(7): 1035-1044.
53. Gruber-Baldini, A. L., et al. (2017). "Validation of the PROMIS measures of self-efficacy for managing chronic conditions." Qual Life Res 26(7): 1915-1924.
54. L. E. Baum and G. R. Sell, "Growth functions for transformations on manifolds," Pac. /. Math., vol. 27, no. 2, pp. 211- 227, 1968. L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat., vol. 41, no. 1, pp. 164-171, 1970.
55. J. Yamato, J. Ohya, K. Ishii, "Recognizing human action in time-sequential images using hidden markov model." In Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92. (pp. 379-385). IEEE.
56. Chung, P. C., & Liu, C. D. (2008). "A daily behavior enabled hidden Markov model for human behavior understanding. Pattern Recognition", 41(5), 1572-1580.
57. J. Corbin, A. Strauss, Grounded theory research: Procedures, canons and evaluative criteria. Zeitschrift für Soziologie 1990, 19(6), 418-427.
58. L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition". Proceedings of the IEEE 1989, 77(2), 257-286.